IML_project

Heikki Nenonen

2022-11-23

Todo

- dummy classifier
- ~~class4 -> event/nonevent, week1 exe~~?
- ~~drop partlybad, pelkkää FALSEa~~
- varianssit mukana/ei mukana? ei one hot -> yksinkertaistaa liikaa ja tarkoitettu kategoriseen dataan
- date? paljon informaatiota, mutta halutaanko muuttujaksi <- opeta 2000-2008, testaa 2009-2011 / kysy slack test_hidden ei - - date, jätetäänkö pois?
- train, test, cv-10?
- itse logisticregression, week2 exe1 <- lasso/ridge
- accuracy, perplexity, week2 exe1
- accuracy of our accuracy? <– malli train+test, vähän parempi kuin pelkkä train?
- class4 -> nonevent/1a/1b/II/
- googlaa mahdollisia malleja

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model

#npf_test = pd.read_csv("initial_data/npf_test_hidden.csv")
npf_train = pd.read_csv("initial_data/npf_train.csv")
```

```
npf = npf_train.set_index("date")
npf = npf.drop(['id', 'partlybad'], axis=1)


class2 = np.array(["nonevent", "event"])
class2 = class2[(npf["class4"]!="nonevent").astype(int)]
npf.insert(loc=0, column="class2", value=class2)

npf
```

```
##               class2    class4  CO2168.mean  ...  UV_B.std   CS.mean    CS.std
## date                                          ...
## 2000-01-17     event        Ib   368.771711  ...  0.018122  0.000243  0.000035
## 2000-02-28  nonevent  nonevent   378.197295  ...  0.003552  0.003658  0.000940
## 2000-03-24     event        Ib   373.043158  ...  0.272472  0.000591  0.000191
## 2000-03-30     event        II   375.643019  ...  0.451830  0.002493  0.000466
## 2000-04-04  nonevent  nonevent   377.661030  ...  0.291457  0.004715  0.000679
## ...              ...       ...          ...  ...       ...       ...       ...
## 2011-08-16  nonevent  nonevent   381.016623  ...  0.496816  0.002423  0.000425
## 2011-08-19  nonevent  nonevent   383.698146  ...  0.726461  0.002476  0.000902
## 2011-08-21  nonevent  nonevent   379.279128  ...  0.363890  0.003484  0.000457
## 2011-08-22  nonevent  nonevent   384.443758  ...  0.595032  0.004782  0.001082
## 2011-08-27  nonevent  nonevent   382.230839  ...  0.722553  0.006956  0.000605
##
## [464 rows x 102 columns]
```